HELMHOLTZ METADATA COLLABORATION

HZB Helmholtz Zentrum Berlin

AST | E&E | ENERGY | HEALTH | INFORMATION | MATTER | DATA COMMONS | HMC PROJECTS | HMC OFFICE

Better data for better science

Helmholtz-Zentrum für Materialien und Energie

Oonagh Mannix

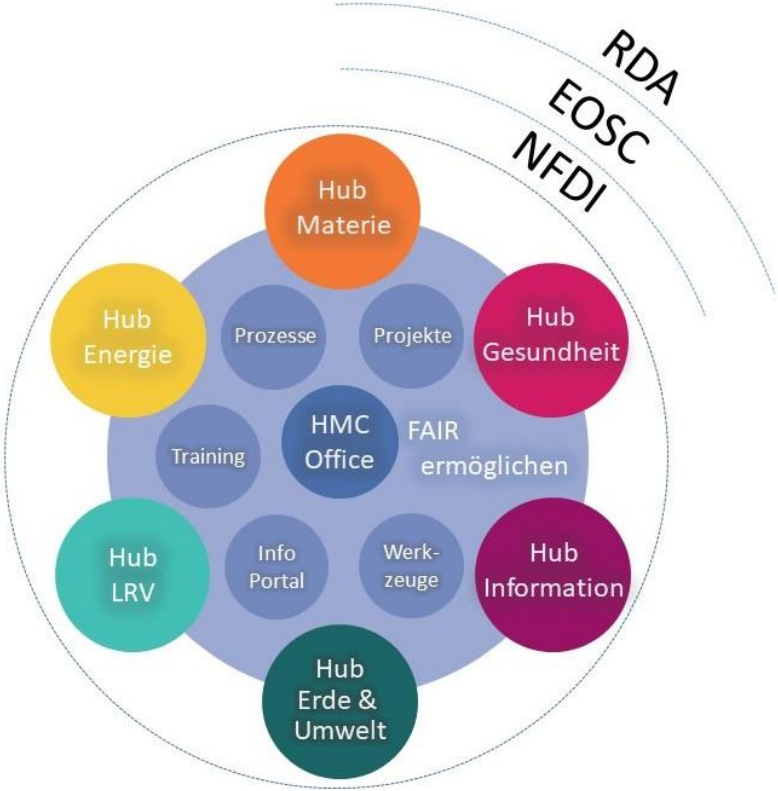
www.helmholtz-metadata.de

Helmholtz Metadata Collaboration

- Make Helmholtz Data **FAIR** - findable, accessible, interoperable and reusable
- Provide a **sustainable** service for efficient metadata handling
- Establish and support a **metadata community** in the respective research areas

HMC Roadmap

- Six **domain specific hubs** to provide advice
- Provide **standards, best practices, processes and tools** for researchers
- Develop **centralized** and **decentralized** metadata services
- Provide education and **training** to ensure behaviour change
- Facilitate community involvement with **community conducted projects**



HMC Locations

HCM Host Centres

- FZJ – Information & FAIR Data
  Forschungszentrum Jülich
- DKFZ – Health
  German Cancer Research Center
- HZB – Matter
  Helmholtz-Zentrum Berlin für Materialien und Energie
- KIT – Energy & FAIR Data
  Karlsruhe Institute of Technology
- DLR – Aeronautics, Space and Transport
  German Aerospace Center
- GEOMAR – Earth & Environment and HMC-Office
  GEOMAR Helmholtz Center for Ocean Research Kiel

HCM Satellites

- HZDR – Energy
  Helmholtz Center Dresden Rossendorf
- AWI – Earth & Environment
  Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung
- GFZ – Earth & Environment
  Helmholtz Center Potsdam German Research Center for Geosciences
- UFZ – Earth & Environment
  Helmholtz Center for Environmental Research

Hub Matter

Research field matter:

- Particle, astroparticle, plasma physics…
- Technology of accelerators, detectors….
- User programs of photon, neutron, laser facilities……

- Survey of data practices across Helmholtz
  - Closes 5th November 2021
- FAIR data assessment for instruments
  - 15th November 14-15h00

- FAIR meets EMIL: Principles in Practice
  - Günther et al. ICALEPCS (2021)
- RDMinfoPool
  - Available on request

- Glossary for electron microscopy community
  - 8th November 14-16h00
- Training course
  - 22-23 November 2021

Better Data -> FAIR data

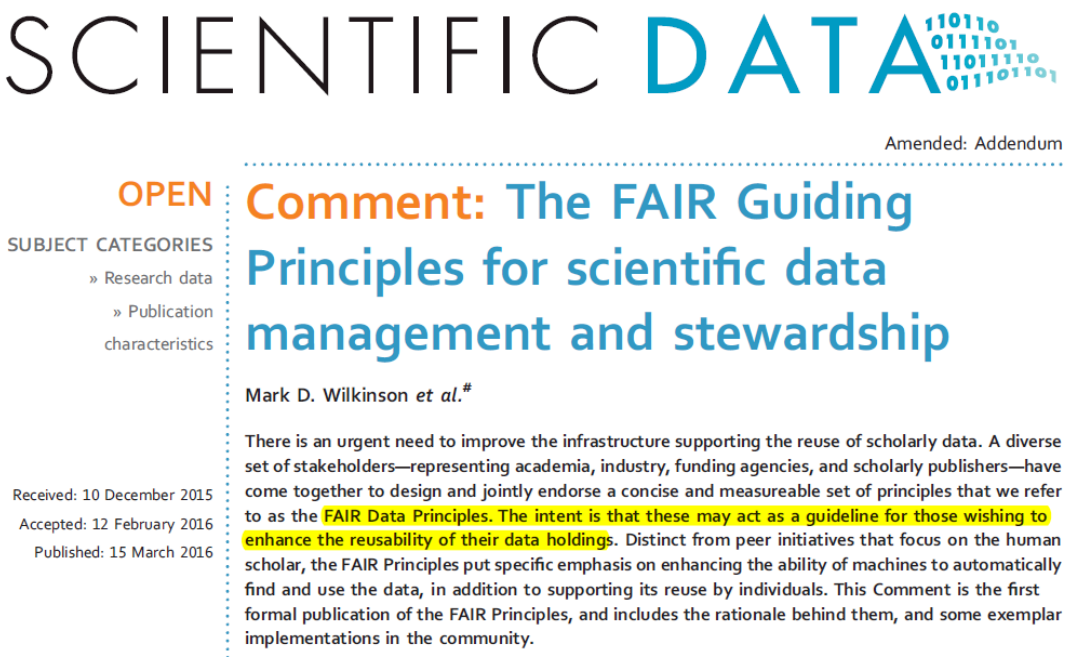
15 FAIR Guiding Principles

F Findable

A Accessible

I Interoperable

R Reusable



SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES

» Research data

» Publication characteristics

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

Comment: The FAIR Guiding Principles for scientific data management and stewardship

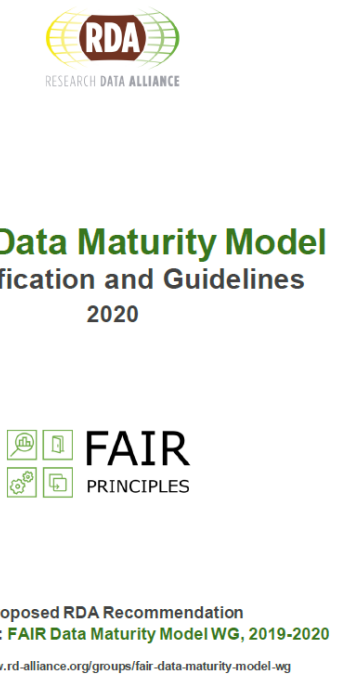
Mark D. Wilkinson *et al.*#

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

FAIR Guiding Principles (2016)

Wilkinson, Mark D., et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific data* 3.1 (2016): 1-9.

RDA: FAIR Data Maturity Model



https://zenodo.org/record/3909563

From 15 FAIR Guidelines (2016) **to 41 FAIR Indicators** (2020)

Example:

Guideline:

F1. (Meta)data are assigned a globally unique and persistent ID

Indicators:

FAIR	ID	Indicator
F1	RDA-F1-01M	Metadata is identified by a persistent identifier
F1	RDA-F1-01D	Data is identified by a persistent identifier
F1	RDA-F1-02M	Metadata is identified by a globally unique identifier
F1	RDA-F1-02D	Data is identified by a globally unique identifier

Rating: **Essential *** – Important ** - Useful ***

Assessing the FAIRness of a prototypical PaN instrument at BESSY II.

Markus Kubin and Gerrit Günther

15th November 14h00-15h00

The authors employed the RDA's FAIR Data Maturity Model to assess the FAIRness of data produced by a prototypical PaN instrument at the BESSY II synchrotron radiation facility and identified steps towards improving the FAIRness.

Ensuring data is FAIR goes beyond an instrument or even an institute:

- Repository
- Format
- Workflow

Better Science

Science that moves from output to outcomes

- What have you discovered?
- Why is it important?
- What have you done about this?

Benedictus et al. Fewer numbers, better science. *Nature* (2016). https://doi.org/10.1038/538453a

What have you discovered?

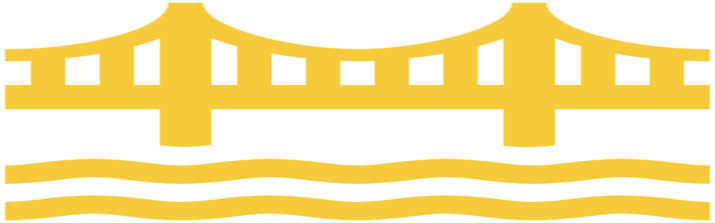
I tried to repeat other researcher's experiment […], but I could not reproduce the data. I felt that the data are misinterpreted and seem suspicious. Therefore, I contacted this researcher but it seems that the researcher is refusing to provide me with the published data. I wish these published data are uploaded on the manuscript upon publication, rather than requesting them to the author.

The only thing that matters and allows a scientist to exist in the job continues to be publication... Any documentation that exceeds this purpose […] is a waste of time. …making data open access means scientific suicide: As own publications are the only thing ultimately maintaining a scientists job, anyone who lets others analyze data without forcing them to involve oneself on a lead position is […] a fool. No scientist receives any budget […]because of documenting metadata systematically or for sharing data open access. This is not good or in the interest of the public that funds all our research, but this is the political design of the system.

HMC annual survey (2021)

Better data requires better incentives for science

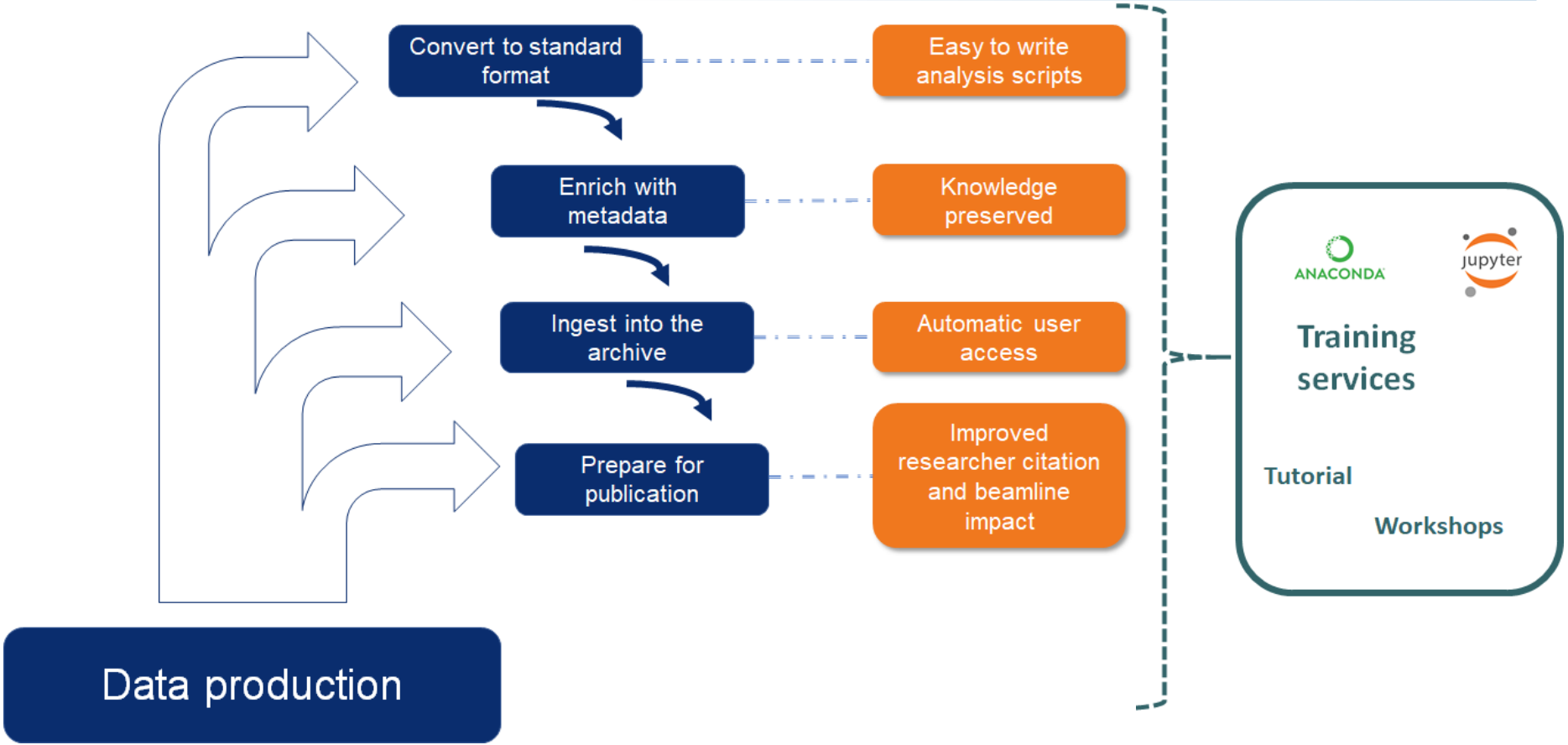
Building bridges



Improve capacity of infrastructure to engage with researchers

Improve capacity of researchers to engage with Infrastructure

Use case: FAIR meets EMIL



Günther et al ICALEPCS 2021

Summary

- FAIR data can be considered better data
  - HMC is working on methods of evaluating this
- Better science requires better incentives
  - Track qualitatively via surveys
- Try to bridge infrastructure and researchers through targeted use cases
- Community engagement is essential

Here to help

- Website:
  - https://helmholtz-metadata.de/en
  - hmchelpdesk@geomar.de
- Twitter
  - @helmholtz_hmc
- Mattermost
  - https://mattermost.hzdr.de/hmc-public
- Open office
  - Every Friday 13h00-16h00
  - https://jitsi.helmholtz-berlin.de/HMC_matter_openoffice
- Email:
  - oonagh.mannix@helmholtz-berlin.de
  - hmc-matter@helmholtz-berlin.de